

저화질 방송미디어의 고품질 변환 기술 개발 현황

□ 조속희, 라상중 / 한국전자통신연구원

요약
과거에 제작된 SD/HD급 미디어에 대한 수요가 증가하고 영상처리 기술 분야에서의 인공지능 적용이 확산되면서 딥러닝 기반 저화질 미디어의 고품질 변환 기술 연구가 활발히 수행되고 있다. 본고에서는 SD/HD급으로 제작된 저화질 미디어를 FHD(Full HD)나 4K-UHD(Ultra HD) 등 고품질 시청 환경에 적합하도록 화질을 개선하여 고품질 미디어로 변환하는데 필요한 주요 요소 기술 및 이를 적용한 고품질 변환 시스템의 개발 현황에 대하여 살펴본다.

I. 서론

미디어의 고품질 변환 기술은 방송미디어 규격의 변화와 미디어 시청 환경의 변화에 의해 시작된 기술분야라 할 수 있다. 먼저 방송미디어 규격의 변화 측면에서 살펴보면, 아날로그 흑백 방송, 컬러 방송을 거쳐

FHD, 4K-UHD 디지털 방송으로 변화하여 왔다. 우리나라 지상파 방송은 1966년에 아날로그 흑백 방송이 전국에 송출되었고, 1981년 컬러 방송이 시작되었으며, 디지털 방송은 2001년 11월에 서울과 수도권을 중심으로 본방송이 시작되어 2006년 7월부터는 전국 방송으로 확대되었다. 2012년 12월 31일부로 아날로그 방송 송출이 중단되기 이전까지 아날로그와 디지털 방송을 동시에 실시하였다. 이 시기에 방송미디어는 Full HD(1920x1080 해상도)가 아닌 SD/HD급 영상이나 아날로그 영상으로 촬영되어 제작되었고, 영상규격의 변환을 통해 아날로그 방송 및 HDTV 방송이 실시되었다. 2017년부터 현재 수도권과 5대광역시를 중심으로 실시되고 있는 4K-UHD 방송콘텐츠의 경우도 마찬가지로 4K 카메라로부터 4K-UHD 규격의 영상을 획득하여 제작/송출되는 경우도 있지만, 일부는 HD 규격으로 촬영

※ 이 원고는 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2021-0-00087, SD/HD급 저화질 미디어의 고품질 변환 기술 개발)

<표 1> 방송 영상 신호 규격

구분	아날로그 방송	디지털 방송		
		SDTV	HDTV	4K-UHDTV
주사선수	525개	-	-	-
해상도	-	704x480/640x480	1920x1080	3840x2160
화면비	4:3	4:3 또는 16:9	16:9	16:9
주사방식	비월주사 (Interlaced scanning)	비월주사 (Interlaced scanning)	비월주사 (Interlaced scanning)	순차주사 (Progressive scanning)
프레임율(fps)	29.97	29.97	29.97	60

된 영상을 4K-UHD 규격으로 변환하여 송출되고 있다. 방송기술이 발전함에 따라 함께 변화되어 온 방송 규격에서의 영상품질은 당연히 고화질화/고선명화를 추구하고 있다. 아날로그 방송, SDTV, HDTV, 4K-UHDTV 방송에 있어서 주요 영상신호 규격은 <표 1>과 같다. 최신 규격으로 변화됨에 따라 해상도 및 프레임율이 높아지면서 고품질화되고 있으며, 기제작된 콘텐츠의 재송출 또는 재촬영을 위해서는 최신 방송 규격에 맞는 콘텐츠 규격으로의 변환이 필요하게 된다. 이때 단순 규격 변환만 이루어지면 영상의 품질저하가 발생하므로 더 좋은 화질로 변환하기 위해서는 고품질 변환 기술 개발이 필요하게 된다.

다음으로 미디어 시청 환경의 변환 측면에서 살펴보면, 디스플레이 기술의 발전에 따라 1960년대 10인치대의 아날로그 TV를 시작으로 현재 100인치 8K-UHD 디스플레이까지 TV 사이즈가 초대형화되고 있는 추세이다. 초대형화되고 있는 TV에서 초고품질의 영상 재생을 통한 실감형/몰입형 미디어를 제공하기 위해서는 고품질 변환 기술의 적용이 필요할 것이다. 더불어 글로벌 한류 열풍의 영향으로 과거에 제작된 구작(舊作) 아날로그 및 SD급 콘텐츠의 해외 수요 또한 증가하면서 이들 구작의 저품질 영상을 고품질 시청 환경에 적합하도록 변환하는 기술에 대한 산업적 수요가 제기되고 있다. 이에 따라, 본고에서는 저화질 미디어를 고품질 미디어로 변환

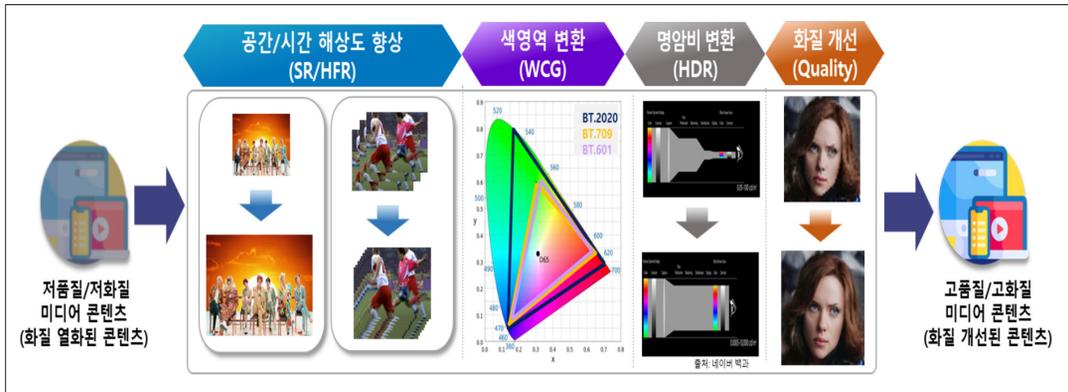
하여 영상규격 요소를 지원하는데 필요한 주요 변환 요소 기술 및 시스템 개발 현황을 살펴본다.

II. 고품질 변환을 위한 주요 기술

저화질 미디어를 고품질화하기 위해서는 <그림 1>에 나타난 바와 같이 초해상화(공간해상도 증가), 고프레임율(시간해상도 증가), 디인터레이싱, 화면비 변환, 명암비 변환, 색영역 변환, 잡음 제거, 디블러링 등의 기술이 적용될 수 있다. 본고에서는 이들 기술 중 주요 기술에 대하여 좀 더 자세히 알아본다.

1. 초해상화(SR: Super Resolution)

초해상화 기술은 영상의 공간해상도를 확장하는 기술로 고품질 변환에 있어 주요 핵심 기술 중 하나이다. 미디어가 아날로그에서 디지털로 전환이 이루어지던 1990년 후반부터 영상신호 처리 기반의 SR 기술 연구가 TV 제조사를 중심으로 수행되어 왔으며, 2010년쯤 UHD 미디어의 등장과 TV 크기의 대형화가 본격적으로 이루어지면서 SR 기술 연구가 다시금 활발히 진행되었다. 영상처리 기반의 SR 기술 성능이 포화상태에 이르던 2015년쯤 인공지능이 활용되기 시작하면서 최근



<그림 1> 고품질 변환 요소

2~3년 사이에 딥러닝 기반 SR 기술과 관련하여 많은 연구 성과가 발표되고 있다.

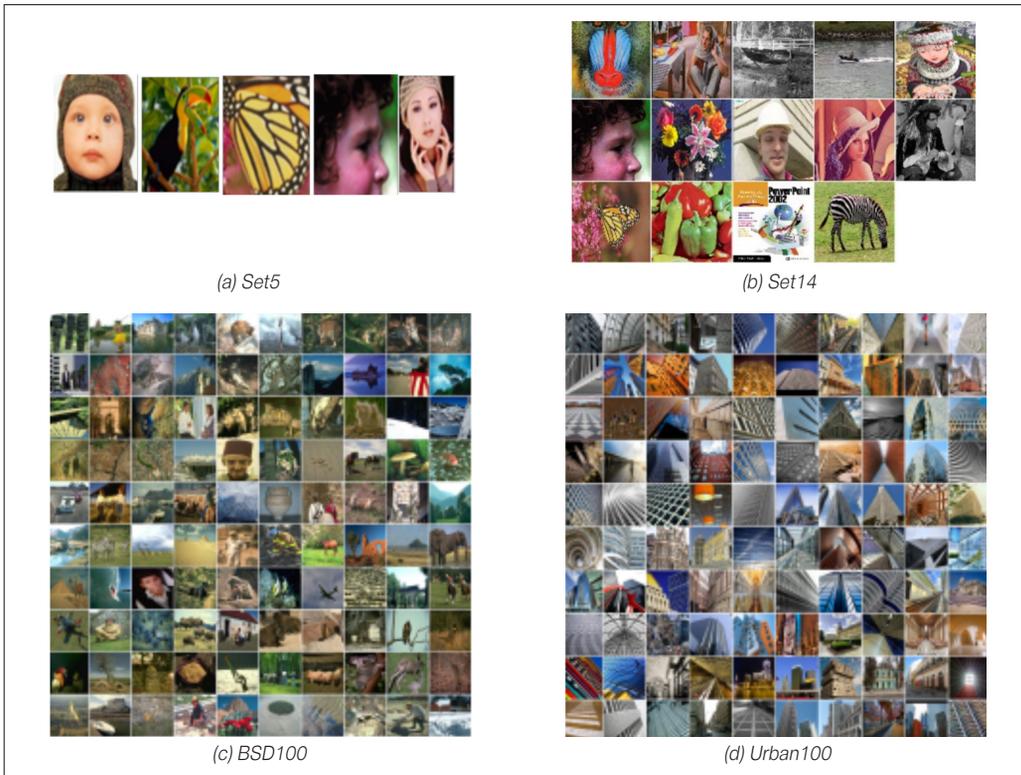
딥러닝 기반 SR 기술 연구는 최근까지 가장 많이 사용되던 CNN(Convolutional Neural Network)이나 RNN(Recurrent Neural network) 모델이 Transformer 신경망 구조로 대체되는 형태로 진행되고 있다. Transformer 모델은 2017년 구글에서 자연어 처리를 위해 제안한 신경망으로, 입력되는 문장 속의 단어들 사이에서 관계를 추적해 맥락과 의미를 학습하는 모델이다[1]. Transformer 모델은 인코더에서 시퀀스를 입력 받고, 디코더에서 시퀀스를 출력하는 인코더-디코더 구조로 구성되어 있다. RNN은 입력 문장에서 단어를 순차적으로 입력하여 처리하면서 각 단어의 위치 정보를 활용하여 이전 단어들을 통해 다음 단어를 학습할 수 있어 시퀀스를 다루는 자연어 처리에 유용하였다. 반면 Transformer 모델은 시퀀스에서 단어를 순차적으로 입력 받지 않는 대신 모든 단어 간 상관관계를 추적한다. 이를 위하여 위치 정보를 나타내는 별도의 방법으로 positional encoding을 수행한다. positional encoding을 통해 입력되는 각각의 단어에 대한 임베딩 벡터에 위치 정보들을 더하여 모델의 입력으로 사용하고 있다.

자연어 처리에서 탁월한 성능이 입증된 Transformer

모델에 대하여 컴퓨터 비전 분야에 처음 적용된 모델이 2021년 발표된 ViT(Vision Transformer)이다[2]. Image Classification을 위해 이미지를 패치 단위로 분할하여 토큰화한 뒤 Transformer 모델에 입력할 수 있도록 한 것이다. ViT 모델이 발표된 이후 다양한 컴퓨터 비전 기술 분야에 적용되고 있으며, 특히, 영상 복원 기술에 적용되면서 SR 기술에서 우수한 성능을 갖는 딥러닝 모델들이 최근 잇따라 발표되고 있다.

SR 기술은 한 장의 저해상도 영상을 이용하여 고해상도 영상으로 변환하는 Image SR 기술과 시간적으로 인접한 복수 장의 영상을 이용하여 고해상도 영상으로 변환하는 Video SR 기술로 구분할 수 있다. Image SR 기술에서는 주로 <그림 2>에 나타낸 Set5, Set14, BSD100, Urban100 데이터셋을 대상으로 객관적 화질 척도인 PSNR(Peak signal-noise ratio)이나 SSIM(Structural similarity index)을 비교하여 성능을 평가하고 있다.

<그림 3>은 Image SR 기술에 있어서 Set5 데이터셋에 대해 4배 확장한 경우에 대한 SOTA(State-Of-The-Art) 기술들을 나타내고 있다. 상위에 랭크되어 있는 HAT(Hybrid Attention)이나 SwinFIR, SwinIR은 Self-attention 메커니즘을 기반으로 하는 Vision Transformer



<그림 2> Image SR에 사용되는 주요 평가 데이터셋

Rank	Model	PSNR ↑	SSIM	MOS	LPIPS	Perceptual Index	Extra Training Data	Paper	Code	Result	Year
1	HAT-L	33.30	0.9083				×	Activating More Pixels in Image Super-Resolution Transformer	Code	Result	2022
2	SwinFIR	33.2	0.9068				×	SwinFIR: Revisiting the SwinIR with Fast Fourier Convolution and Improved Training for Image Super-Resolution	Code	Result	2022
3	HAT	33.18	0.9073				×	Activating More Pixels in Image Super-Resolution Transformer	Code	Result	2022
4	SwinIR	32.93	0.9043				×	SwinIR: Image Restoration Using Swin Transformer	Code	Result	2021
5	LTE	32.81					×	Local Texture Estimator for Implicit Representation Function	Code	Result	2021
6	HAN+	32.75	0.9016				×	Single Image Super-Resolution via a Holistic Attention Network	Code	Result	2020
7	PFF	32.74	0.9021				×	Image Reconstruction with Predictive Filter Flow	Code	Result	2018
8	DRLN+	32.74	0.9013				×	Densely Residual Laplacian Super-Resolution	Code	Result	2019
9	SRGAN + Residual-in-Residual Dense Block	32.73	0.9011				×	ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks	Code	Result	2018
10	RFN	32.71					×	Progressive Perception-Oriented Network for Single Image Super-Resolution	Code	Result	2019

<그림 3> Image SR에서 현재 SOTA 기술 순위 (Set5 데이터셋 대상 결과)

<표 2> Video SR에서 사용되는 주요 학습/평가 데이터셋 현황

데이터셋명	연도	유형	Video 수	해상도	색공간
UVG	2020	Test	16	3,840 x 2,160	YUV
REDS	2019	Train+Test	270	1,920 x 1,080	RGB
Vimeo90K	2019	Train+Test	91,701	448 x 256	RGB
CDVL	2016	Train	-	1,920 x 1,080	RGB
Myanmar	2014	Train	1	3,840 x 2,160	RGB
YUV21	2014	Test	21	352 x 288	YUV
Vid4	2011	Test	4	720 x 480	RGB

신경망 구조를 사용하고 있다. 이들 모델들은 Set5 데이터셋 뿐만 아니라 Set14, Urban100 등의 데이터셋에 대해서도 우수한 성능 결과를 보여주고 있다[3].

Video SR 기술에서는 <표 2>에 나타난 데이터셋이 학습이나 평가에 사용되고 있으며, 평가척도는 Image SR에서와 마찬가지로 PSNR과 SSIM으로 비교하고 있다[4]. <그림 4>는 Video SR 기술에 있어서 Vid4 데이터셋에 대해 4배 확장한 경우에 대해 상위 성능의 기술을 나타내고 있다[5]. Image SR 기술에서와 마찬가지로

Transformer를 기반으로 하는 PSRT 및 RVRT 모델이 우수한 성능을 보여주고 있음을 확인할 수 있다.

딥러닝 기반의 기술 연구에서 우수한 성능의 결과물을 도출하기 위해서는 신경망 구조가 중요하지만 학습에 이용되는 데이터셋 구축 또한 중요한 요소이다. 딥러닝 기반 SR 기술 연구에서도 마찬가지로 학습 데이터셋 구축이 중요하다. 딥러닝 기반 SR 연구를 위해서는 저화질의 저해상도 영상과 고품질의 고품해상도 영상이 쌍(pair)으로 된 데이터셋이 필요하지만 이러한 형태의 데

Rank	Model	SSIM ↑	PSNR	MOVIE	Paper	Code	Result	Year
1	PSRT	0.8485	28.07		Rethinking Alignment in Video Super-Resolution Transformers	Code	Result	2022
2	RVRT	0.8462	27.99		Recurrent Video Restoration Transformer with Guided Deformable Attention	Code	Result	2022
3	iSeeBetter	0.835	27.43		iSeeBetter: Spatio-Temporal Video Super Resolution using Recurrent-Generative Back-Projection Networks	Code	Result	2020
4	VSR-DUF	0.832	27.31		Deep Video Super-Resolution Network Using Dynamic Upsampling Filters Without Explicit Motion Compensation	Code	Result	2018
5	EDVR	0.8264	27.35		EDVR: Video Restoration with Enhanced Deformable Convolutional Networks	Code	Result	2019
6	FRVSR	0.822	26.69		Frame-Recurrent Video Super-Resolution		Result	2018
7	RRN-L	0.8209	27.69		Revisiting Temporal Modeling for Video Super-resolution	Code	Result	2020
8	RBPN/6-PF	0.818	27.12		Recurrent Back-Projection Network for Video Super-Resolution	Code	Result	2019
9	DRDVSR	0.774	25.88		Detail-revealing Deep Video Super-resolution	Code	Result	2017
10	SOF-VSR	0.772	26		Deep Video Super-Resolution using HR Optical Flow Estimation	Code	Result	2020

<그림 4> Video SR에서 현재 SOTA 기술 순위 (Vid4 데이터셋 대상 결과)

이더셋을 카메라 촬영을 통해 획득하는 것은 불가능할 것이다. 동일한 환경에서 동일한 장면의 저해상도 영상과 고해상도 영상을 획득하기 위해서는 저해상도 카메라와 고해상도 카메라를 동일한 위치에서 설치하고 동시간에 획득해야 하는데 이것은 불가능하기 때문이다. 이에 따라 대부분의 SR 기술 연구에서는 고해상도 영상에서 화질열화 및 해상도 축소를 통해 저해상도 영상을 생성하여 학습 데이터셋으로 사용하고 있다. 이렇게 만들어진 저화질 영상은 실제 카메라에서 획득된 저화질 영상의 특성과는 차이가 존재한다. 특히 고품질의 4K나 2K 해상도 영상을 화질열화하여 만든 SD급 해상도의 영상은 실제 카메라에서 획득된 SD급 영상과는 화질열화 특성이 확연히 다르다.

현재 우수한 성능을 보이고 있는 딥러닝 기반 SR 기술을 적용하여 FHD 콘텐츠를 가로세로 각각 2배 확대한 4K-UHD 콘텐츠로의 변환은 어느 정도 고품질 영상으로 변환이 가능한 수준이다. 하지만 저화질의 SD 콘텐츠를 가로세로 2배하여 변환되는 FHD 콘텐츠는 실제 FHD 카메라로 촬영된 콘텐츠와는 화질 차이가 크다. 또한 SR 기술에 의해 변환된 영상을 재생하는 디스플레이 크기, 시청자 개인 성향, 원본 영상의 화질열화 정도 등에 따라 시청자의 주관적 화질 만족도는 다르기 때문에 성능 향상을 위한 SR 기술 연구는 계속 이루어져야 할 것이다.

최근 케이블TV, IPTV, 유튜브 등에서는 오래전에 제작된 다양한 종류의 명작 K-드라마가 재방영되고 있으며, 시청률 또한 유의미한 수준으로 집계되고 있다. 콘텐츠 제작 시기가 오래된 것일수록 화질열화 정도는 더욱 심하게 나타나는 것이 현실인데, 한류 콘텐츠 수요 증가 및 명작 K-드라마 재활용 등의 측면에서 보면 저화질의 SD급 영상을 2배~4배하여 FHD나 4K-UHD 영상으로 변환하였을 때 시청자들이 불편함 없이 즐길 수 있는 화질의 영상으로 변환할 수 있는 SR 기술 연구가 반드시 필요할 것이다.

2. 디인터레이싱(De-Interlacing)과 고프레임율(HFR: High Frame Rate)

〈표 1〉에 나타난 방송 영상 신호 규격에서 알 수 있듯이 아날로그 방송부터 HDTV 방송까지는 한 프레임의 영상을 홀수와 짝수 줄(line)로 나눈 odd-field/even-field를 번갈아 가며 표시하는 비월주사 방식으로 초당 59.94필드(초당 29.97프레임)를 표시한다. UHDTV 방송 규격에서는 한 프레임 전체를 온전하게 표시하는 순차주사 방식을 사용하여 초당 60프레임으로 동작한다. 흔히 60i로 표기되는 프레임율은 초당 59.94회 인터레이스 스캔을 의미하며, 60p, 30p로 표기되는 프레임율은 각각 초당 59.94회, 29.97회 프로그레시브 스캔을 의미한다. 이에 60i와 60p 사이의 프레임율은 2배 차이가 존재하게 되므로 60i 영상을 디인터레이싱하여 60p로 변환하면 2배 확장된 프레임율이 된다.

현재 4K-UHD 방송에서는 60p 프레임율을 지원하지만 공간해상도가 높아질수록 프레임율 또한 함께 높아져야 고품질의 영상을 제공할 수 있다. 이에 8K-UHD, 16K-UHD에서는 120p, 240p 등의 고프레임율이 거론되고 있다. 프레임율은 움직임이 빠른 축구, 농구 등 스포츠 영상에서 품질의 차이를 확연히 느낄 수 있다. 예를 들어 골프 영상에서 날아가는 골프공의 모양을 제대로 재생하기 위해서는 공간해상도 뿐만 아니라 시간해상도인 프레임율의 향상이 수반되어야 할 것이다.

HFR 변환 기술은 크게 optical-flow 기반 방식과 kernel-regression 기반 방식으로 구분할 수 있다. optical-flow 기반 방식은 일반적으로 선형 또는 2차 모델을 기반으로 이미지나 특징을 왜곡시킨 후 왜곡된 결과를 융합함으로써 중간 프레임을 생성하게 된다. 움직임이 복잡한 경우 실제 모션을 모델링하기에는 충분하지 않은 편이다. 반면 kernel-regression 기반 방식은 암시적 방식으로 추론을 수행하며, 중간 프레임의 픽셀

Rank	Model	PSNR ↑	SSIM	LPIPS	Speed (ms/f)	Paper	Code	Result	Year
1	MA-CSPA	36.76	0.9800			Exploring Motion Ambiguity and Alignment for High-Quality Video Frame Interpolation		📄	2022
2	IFRNet	36.20	0.9808		16 (Tesla V100)	IFRNet: Intermediate Feature Refine Network for Efficient Frame Interpolation	🔗	📄	2022
3	ABME	36.18	0.9805			Asymmetric Bilateral Motion Estimation for Video Frame Interpolation	🔗	📄	2021
4	RIFE_Large	36.10	0.9800			RIFE: Real-Time Intermediate Flow Estimation for Video Frame Interpolation	🔗	📄	2020
5	SoftSplat	36.10	0.970			Softmax Splatting for Video Frame Interpolation	🔗	📄	2020
6	NCM-Base	35.88	0.9795			Neighbor Correspondence Matching for Flow-based Video Frame Synthesis		📄	2022
7	CURE	35.73	0.9789			Learning Cross-Video Neural Representations for High-Quality Frame Interpolation	🔗	📄	2022
8	M2M-PWC	35.4	0.978		32 (Titan X)	Many-to-many Splatting for Efficient Video Frame Interpolation	🔗	📄	2022
9	RRIN	35.22	0.9643			Video Frame Interpolation via Residue Refinement	🔗	📄	2020
10	CDFI	35.17		0.010		CDFI: Compression-Driven Network Design for Frame Interpolation	🔗	📄	2021

<그림 5> HFR 변환 기술에서 현재 SOTA 기술 순위 (Vimeo90K 데이터셋 대상 결과)

을 생성하기 위해서 이미지나 특징에서 인접한 픽셀을 적용적으로 계산하여 생성한다. 하지만 이 방식은 고�해상도 영상이나 움직임이 큰 영상에서의 성능이 떨어지는 문제점이 있다.

<그림 5>는 HFR 변환 기술에 있어서 Vimeo90K 데이터셋에 대한 SOTA 기술들을 나타내고 있다[6]. <그림 5>에서 알 수 있듯이 상위에 랭크되어 있는 HFR 변환 기술들은 아직 Transformer 모델을 적용하지 않고 있지만, 최근 CVPR2022에 Transformer 모델을 적용한 논문이 발표되었다[7]. 본 논문에서 제안하는 모델의 성능은 Vimeo90K 데이터셋에 대해 PSNR 36.5dB, SSIM 0.9816으로 <그림 5>에 나타난 상위 랭크 모델과 거의 유사한 성능을 보이고 있다.

SR 기술과 비교하여 HFR 변환 기술은 학술적으로는 높은 성능을 보여주고 있지만, 실환경에서 유통되고 있는 방송미디어 변환에 적용할 경우 해결해야 하는 추가적인 문제가 있다. 일반적으로 콘텐츠가 제작되기까지는 카메라-저장장치-편집의 단계를 거치게 되는데, 이 3

단계에서 프레임율의 변환이 많이 발생하고 있다. 특히 방송미디어의 경우는 카메라에서 24p로 촬영하여 저장 장치나 편집과정을 거치면서 60i로 변환되어 제작되는 경우가 많으며, 이때 odd-field/even-field가 거꾸로 저장되는 오류도 많이 발생하고 있다. 이렇게 촬영본과 최종 편집본(송출본)에서의 프레임율이 변환되어 있는 인터레이스 영상을 프로그레시브 영상으로 변환하게 되면 객체의 움직임이 부자연스럽게 되는 문제를 야기시킨다. 이에 실환경에서 유통되고 있는 방송미디어의 디인터레이싱이나 고프레임율 변환을 위해서는 원본 영상의 촬영 당시 프레임율을 파악하는 것이 우선적으로 고려되어야 우수한 품질의 결과를 얻을 수 있을 것이다.

3. 화면비(Aspect Ratio)

동영상 플랫폼에 따른 화면비는 <표 3>과 같이 다양하다. 영화 콘텐츠가 TV에서 방영될 때 화면 상하단에 black-bar가 있는 이유가 화면비 때문이다. 방송미디어

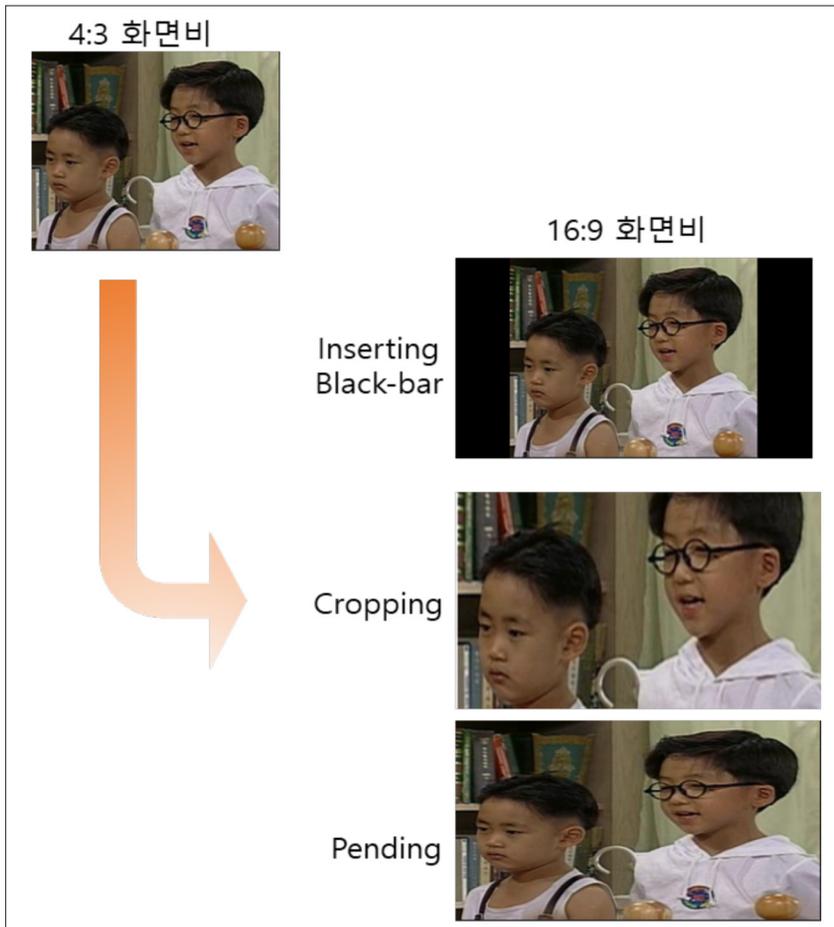
<표 3> 동영상 플랫폼에 따른 화면비

동영상 플랫폼	화면비
SDTV	4:3(1.33:1)
와이드 SDTV / HDTV / 4K-UHDTV	16:9(1.78:1)
디지털 영화	1.85:1
표준 와이드 스크린 영화	21.3:9(2.37:1)
35mm 필름 / 디지털 SLR 카메라	3:2(1.5:1)

에서의 화면비 변환은 4:3을 16:9 비율로의 변환이 필요하게 되는데, SD급 미디어의 FHD 미디어로의 변환에 있어 가장 어려운 변환 기술이라 할 수 있다. 4:3을 16:9

로 변환하는 기술은 SR 기술이나 HFR 변환 기술 대비 수요가 많지 않아 연구가 활발히 진행되고 있지 않다. SDTV에서 HDTV로 전환되던 2000년도 초반은 <그림 6>에 나타낸 바와 같이 화면 좌우측에 black-bar를 삽입하거나, 화면 상하단을 잘라내거나, 가로방향으로 늘리는 방식으로 16:9 화면비의 영상을 만들었다. 어느 방식이든 시청시에 어색함이나 불편함을 느끼게 하므로 SD급 미디어의 재활용을 위해서는 반드시 개선 연구가 필요한 변환 요소라 할 수 있다.

화면비 변환 기술은 SR 기술을 이용하여 가로 해상도만 향상시키거나 객체와 배경을 구분하여 배경 화

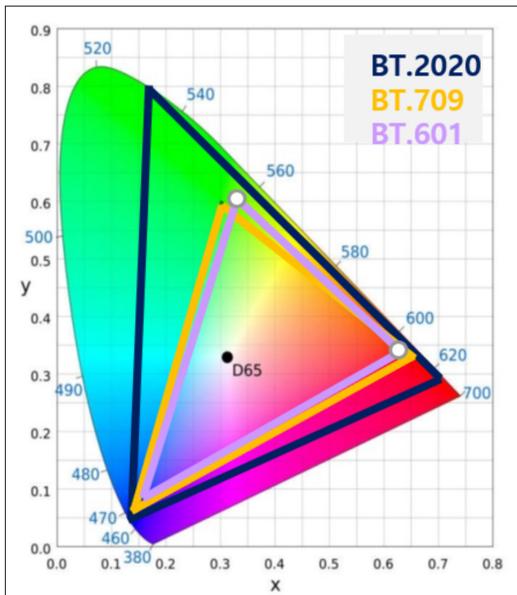


<그림 6> 4:3 to 16:9 화면비 변환 기존 변환 방식

면에 대해 가로 해상도만 향상시키고 객체를 삽입하는 방식 등이 고려될 수 있다. 물론 GAN(Generative adversarial network)을 이용하여 좌우측의 일부 영상을 인위적으로 생성하여 16:9 영상으로 변환하는 방식도 고려할 수 있지만, 데이터셋이 없어 연구에 어려움이 있다. 이렇듯 화면비 변환 기술은 SR 기술, In-painting 기술, video retargeting 기술 등을 이용하여 객체들이 왜곡된 비율로 보이지 않도록 변환하는 방식의 연구가 이루어져야 할 것이다.

4. 색영역(WCG)과 명암비(HDR) 변환

영상이 고품질화되면서 색영역과 명암비 또한 확장되었다. SD급, FHD, UHD 미디어의 색영역은 각각 BT.601, BT.709, BT.2020의 규격으로 <그림 7>에 나타난 크기의 색영역을 표시할 수 있으며 신호 처리에 의해 확장 변환할 수 있다[8-10]. 마찬가지로 명암비 확장



<그림 7> SDTV, HDTV, UHDTV 색영역

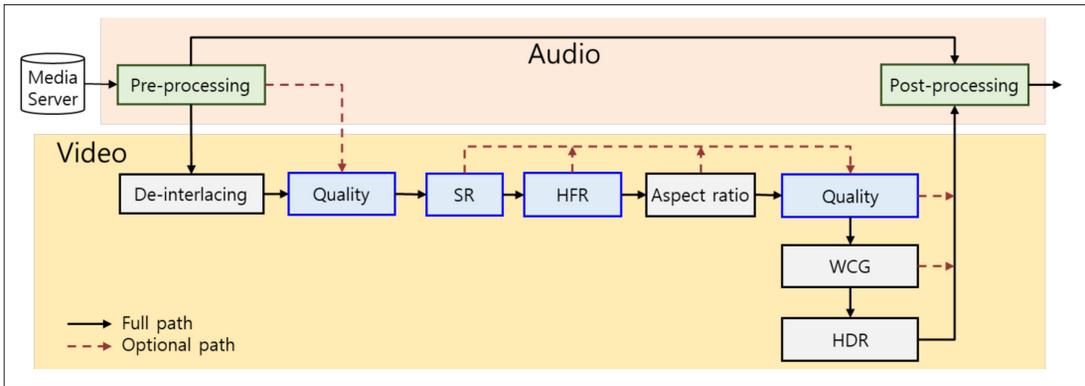
은 Dolby에서 개발한 SMPTE ST 2084/ARIB STD-B67 표준을 기반으로 신호처리에 의해 SDR(Standard Dynamic Range)를 HDR(High Dynamic Range)로 변환할 수 있다[11-12].

색영역과 명암비 변환은 미디어 제작자의 주관적 의견이 반영되는 감각적인 내용이 반영되는 변환 요소이며 콘텐츠 저작권과도 연관될 수 있어 획일적인 변환을 수행하는 데는 기술적 이외의 한계가 존재하는 변환 요소이다.

III. 고품질 변환을 위한 시스템 기술

SD급 저화질 미디어를 고품질의 FHD, UHD 미디어로 재탄생시키기 위해서는 <그림 8>에 나타난 일레와 같이 2장에서 살펴본 주요 변환 요소 기술들이 복합적으로 수행되어야 한다.

저화질 미디어의 고품질 변환에 있어서 SR이나 HFR은 반드시 수행해야 하는 변환 요소이지만 입력되는 영상규격이나 제작자의 의도에 따라 디인터레이싱, 화질 개선(Quality), 색영역 확장(WCG), 명암비 확장(HDR), 화면비 변환은 선택변환될 수 있다. 또한 고품질 변환을 품질에 우선할지, 실시간 처리에 우선할지 등 사업화 모델에 따라 고품질 변환 시스템 개발 방향은 달라질 것이다. 현재는 변환 요소 기술들을 통합 적용한 사업화 모델 기반의 고품질 변환 시스템이나 솔루션 개발은 미비한 상태이다. <표 4>에 나타난 고품질 변환 시스템 개발 현황 및 이에 따른 개발 고려 사항을 분석해볼 때, 고품질 변환 시스템의 상용화를 위해서는 <그림 9>와 같이 고품질 변환 요소 기술의 성능 고도화와 더불어 이들 변환 요소를 복합적으로 실시간 및 비실시간으로 변환할 수 있는 시스템 기술 개발이 요구된다.



<그림 8> 저화질 미디어의 고품질 변환 요소별 적용 흐름도 일례

<표 4> 고품질 변환 시스템 개발 현황 및 개발 고려 사항

현황	개발 고려 사항
영상 변환 요소별 Tool 제공 솔루션은 상용 방송미디어 변환 적용에 어려움 존재	방송/OTT에서 유통되는 AV 미디어 규격을 지원하는 End-to-end 변환 솔루션 필요
방송사/콘텐츠 제작사/개인미디어 제작자 등 다양한 수요자를 위한 네트워크 환경에서의 변환 솔루션 미비	클라우드 환경 지원
콘텐츠 영상 특성이나 제작자 의도를 반영할 수 있는 시스템/솔루션 부재	비실시간/대화형 편집형 변환 지원
스마트폰, 태블릿, 대형TV 등의 콘텐츠 소비환경에 적절한 품질의 미디어로 변환하는 서비스 솔루션 미비	실시간/고품질 변환 미디어 스트리밍 지원



<그림 9> 고품질 변환 요소 기술을 적용한 고품질 변환 시스템/솔루션

<표 5> 주요 변환 요소별 변환 범위 및 내용

주요 변환 요소	변환 범위 및 내용
시공간해상도 변환	- 인터레이스 → 프로그레시브 변환 - SD(480@59.94i) → FHD(1080@60p)/4K-UHD(2160@120p) 변환 - HD(720@30p) → FHD(1080@60p)/4K-UHD(2160@120p) 변환 - FHD(1080@59.94i) → 4K-UHD(2160@120p)/8K-UHD(4320@240p) 변환 - 4K-UHD(2160@60p) → 8K-UHD(4320@120p) 변환 - 4K-UHD(2160@120p) → 8K-UHD(4320@240p) 변환
색영역 변환	- BT.601 → BT.709 - BT.709 → BT.2020
명암비 변환	- 8bits/pixel, Gamma-TF → 10bits/pixel, HLG
화질 개선	- De-blurring, 압축왜곡 개선, 화면비 변환

현재 유통되고 있는 주요 미디어 규격을 고려할 때 고품질 변환 시스템에서 검토될 수 있는 고품질 변환 요소별 변환 범위 및 내용은 <표 5> 정도일 것이다. <표 5>에 나타난 변환 요소들이 영상의 특성을 반영한 적응적 고품질 변환을 지원하고, 변환 요소들이 모든 영상에 일률적으로 동일하게 적용되는 것이 아니라 사용자 의도에 따라 변환하고자 하는 요소들을 동적으로 선택하여 적용할 수 있는 동적 대화형 고품질 변환을 지원하는 형태의 시스템 개발이 필요할 것으로 생각된다.

IV. 결론

본고에서는 SD/HD급의 저화질 미디어를 FHD나 4K-UHD의 고품질 미디어로의 변환에 필요한 주요 변환 요소 기술 개발 현황과 시스템 개발에 있어서의 고려 사항을 기술하였다. 글로벌 한류 열풍의 영향으로 수요가 증가하고 있는 구작(舊作) 아날로그 및 SD급 미디어가 고품질 미디어로 재탄생할 수 있도록 우수한 성능의 변환 요소 기술과 시스템 기술이 연구개발되기를 기대한다.

참고 문헌

- [1] "Attention is all you need," Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin, NIPS2017
- [2] "An image is worth 16x16 words: transformers image recognition at scale," Alexey Dosovitskiy et al., ICLR 2021
- [3] <https://paperswithcode.com/sota/image-super-resolution-on-set5-4x-upscaling>
- [4] <http://ainote.ddlee.cc/2020/12/26/video-super-resolution-datasets/>
- [5] <https://paperswithcode.com/sota/video-super-resolution-on-vid4-4x-upscaling>
- [6] <http://paperswithcode.com/sota/video-frame-interpolation-on-vimeo90k>
- [7] "Video frame interpolation with transformer," Liying Lu et al.,
- [8] "Recommendation ITU-R BT.601-7: Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios," ITU-R, 2011.03
- [9] "Recommendation ITU-R BT.709-6: Parameter values for the HDTV standards for production and international programme exchange," ITU-R, 2015.06
- [10] "Recommendation ITU-R BT.2020-2: Parameter values for the HDTV standards for production and international programme exchange," ITU-R, 2015.10
- [11] "SMPTE ST2084: High dynamic range electro-optical transfer function of mastering reference displays," the Society of Motion Picture and Television Engineers Journal, pp. 1-14, 2014
- [12] "ARIB STD-B67: Parameter values for the hybrid log-gamma (HLG) high dynamic range television (HDR-TV) system for programme production," ARIB, 2015

필자 소개



조 숙 희

- 1993년 2월 : 부경대학교 전자계산학과 이학사
- 1995년 2월 : 부경대학교 전자계산학과 이학석사
- 1999년 9월 : 요코하마국립대학교 전자정보통신공학과 공학박사
- 1999년 11월 ~ 현재 : 한국전자통신연구원 책임연구원
- 2020년 6월 ~ 현재 : TTA PG804 부의장
- 2022년 1월 ~ 현재 : 미래방송미디어표준포럼 UHD융합기술 분과위원회 의장
- 주관심분야 : 영상처리, UHDTV 방송 기술



라 상 중

- 2002년 2월 : 전남대학교 컴퓨터공학과 공학사
- 2004년 2월 : 전남대학교 컴퓨터정보통신공학과 공학석사
- 2018년 2월 : 전남대학교 전자컴퓨터공학과 공학박사
- 2021년 3월 ~ 현재 : 한국전자통신연구원 책임연구원
- 주관심분야 : 영상처리, UHDTV 방송 기술